



Benchmarking framework for myocardial tracking and deformation algorithms: an open access database

Catalina Tobon-Gomez, Mathieu de Craene, Kristin Mcleod, Lennart Tautz, Wenzhe Shi, Anja Hennemuth, Adityo Prakosa, Hengui Wang, Gerald Carr-White, Sergio Kapetanakis, et al.

► To cite this version:

Catalina Tobon-Gomez, Mathieu de Craene, Kristin Mcleod, Lennart Tautz, Wenzhe Shi, et al.. Benchmarking framework for myocardial tracking and deformation algorithms: an open access database. Medical Image Analysis, 2013, 17 (6), pp.632-648. 10.1016/j.media.2013.03.008 . hal-00855928

HAL Id: hal-00855928

<https://inria.hal.science/hal-00855928>

Submitted on 6 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Benchmarking Framework for Myocardial Tracking and Deformation Algorithms: An Open Access Database

C. Tobon-Gomez^{a,g,1,1,*}, M. De Craene^{a,b,1,**}, K. McLeod^{c,2}, L. Tautz^{d,2}, W. Shi^{e,2}, A. Hennemuth^{d,3}, A. Prakosa^{c,3}, H. Wang^{e,3}, G. Carr-White^{1,4}, S. Kapetanakis^{1,4}, A. Lutz^{f,4}, V. Rasche⁴, T. Schaeffter^{1,4}, C. Butakoff^{a,g,5}, O. Friman^{i,5}, T. Mansi^{h,5}, M. Sermesant^{c,5}, X. Zhuang^{j,5}, S. Ourselin^{j,6}, H-O. Peitgen^{d,6}, X. Pennec^{c,6}, R. Razavi^{1,6}, D. Rueckert^{e,6}, A. F. Frangi^{a,k,7}, K. S. Rhode^{1,7}

^a*CISTIB, Universitat Pompeu Fabra, Barcelona, Spain*

Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, Barcelona, Spain

^b*Philips Research, Medisys, Suresnes, France*

^c*Inria-Asclepios Project, Sophia Antipolis, France*

^d*Fraunhofer MEVIS, Bremen, Germany*

^e*Imperial College London, London, United Kingdom*

^f*Department of Internal Medicine II - Cardiology, University of Ulm, Ulm, Germany*

^g*PhySense, Universitat Pompeu Fabra, Barcelona, Spain*

^h*Siemens Corporation, Corporate Research and Technology, Princeton, NJ, U.S.A*

ⁱ*Swedish Defence Research Agency, Linköping, Sweden*

^j*University College London, London, United Kingdom*

^k*Department of Mechanical Engineering, University of Sheffield, Sheffield, United Kingdom*

^l*Division of Imaging Sciences & Biomedical Engineering, King's College London, London, United Kingdom*

Abstract

In this paper we present a benchmarking framework for the validation of cardiac motion analysis algorithms. The reported methods are the response to an open challenge that was put to the medical imaging community through a MICCAI workshop. The database included magnetic resonance (MR) and 3D ultrasound (3DUS) datasets from a dynamic phantom and 15 healthy volunteers. Participants processed 3D tagged MR datasets (3DTAG), cine steady state free precession MR datasets (SSFP) and 3DUS datasets, amounting to 1158 image volumes. Ground-truth for motion tracking was based on 12 landmarks (4 walls at 3 ventricular levels). They were manually tracked by two observers in the 3DTAG data over the whole cardiac cycle, using an in-house application with 4D visualization capabilities. The median of the inter-observer variability was computed for the phantom dataset (0.77mm) and for the volunteer datasets (0.84mm). The ground-truth was registered to 3DUS coordinates using a point based similarity transform. Four institutions responded to the challenge by providing motion estimates for the data: Fraunhofer MEVIS (MEVIS), Bremen, Germany; Imperial College London - University College London (UCL), UK; Universitat Pompeu Fabra (UPF), Barcelona, Spain; Inria-Asclepios project (INRIA), France. Details on the implementation and evaluation of the four methodologies are presented in this manuscript. The manually tracked landmarks were used to evaluate tracking accuracy of all methodologies. For 3DTAG, median values were computed over all time frames for the phantom dataset (MEVIS=1.20mm, IUCL=0.73mm, UPF=1.10mm, INRIA=1.09mm) and for the volunteer datasets (MEVIS=1.33mm, IUCL=1.52mm, UPF=1.09mm, INRIA=1.32mm). For 3DUS, median values were computed at end diastole and end systole for the phantom dataset (MEVIS=4.40mm, UPF=3.48mm, INRIA=4.78mm) and for the volunteer datasets (MEVIS=3.51mm, UPF=3.71mm, INRIA=4.07mm). For SSFP, median values were computed at end diastole and end systole for the phantom dataset (UPF=6.18mm, INRIA=3.93mm) and for the volunteer datasets (UPF=3.09mm, INRIA=4.78mm). Finally, strain curves were generated and qualitatively compared. Good agreement was found between the different modalities and methodologies, except for radial strain that showed a high variability in cases of lower image quality.

Keywords: Cardiac Motion Tracking, 3D tagged MRI, 3D ultrasound, Multimodal, Spatiotemporal registration

*St. Thomas Hospital, Westminster Bridge Road, London, UK

**33 Rue de Verdun, 92156 Suresnes, France

Email addresses: catalina.tobon_gomez@kcl.ac.uk

(C. Tobon-Gomez), mathieu.de_craene@philips.com

(M. De Craene)

¹C. Tobon-Gomez and M. De Craene share lead coauthorship.

²Lead author in challenge submission.

³Joint lead author in challenge submission.

⁴Volunteer and phantom dataset acquisition.

⁵Code, algorithm support and cosupervision.

⁶Group supervision.

⁷A. F. Frangi and K. S. Rhode share senior coauthorship.

1. Introduction

Despite the rapid development of analysis tools in medical imaging, their translation to the clinical environment is limited. One of the main limitations for algorithms to be adopted in the clinical arena is the lack of proper validation strategies. Even for algorithms with extensive in-house validation, a comparison with *state-of-the-art* techniques is difficult due to differences in the evaluated datasets (i.e. type of patients, number of cardiac phases, image quality, image resolution).

The benchmarking of algorithms is a very important activity as we move from bench to bedside in the medical image processing community. In the last few years, several conferences in the Medical Image Analysis field, have provided the opportunity to benchmark algorithms from multiple research groups. In the context of a *challenge*, the organizers provide datasets to test the algorithms. Several researchers process the datasets with their algorithms and submit their results to be evaluated in a unified manner. For an updated list of these challenges, please visit <http://www.grand-challenge.org/>.

In this paper we present a benchmarking framework for the validation of cardiac motion analysis algorithms. The reported methods are the response to an open challenge that was put to the medical imaging community through a MICCAI workshop (Tobon-Gomez et al., 2012). Each participant quantified myocardial motion in the left ventricle from cine and/or tagged Magnetic Resonance (MR) and/or 3D ultrasound (3DUS) modalities. Four institutions responded to the challenge. The datasets and ground-truth used in this framework are publicly available via the Cardiac Atlas Project⁸. Hereby we present a *state-of-the-art* reference multimodal database for the validation of current and future cardiac motion analysis algorithms.

2. Data acquisition

2.1. Database

For any modality, validation of myocardial motion and deformation quantification algorithms is a challenging task that requires reference databases with controlled ground-truth. Unfortunately, in the medical field, it is unfeasible to cover the whole spectrum between controlled ground-truth and data realism. As you increase one, the other one is reduced. For this reason, this challenge focused on a subpart of the spectrum including two types of data: phantom and volunteer datasets. Acquiring images from a physical phantom, we reduce acquisition artifacts hence providing a higher quality ground-truth. Acquiring images from young volunteers, we increase the realism of the datasets while representing healthy cardiac deformation (Wandt et al., 1998; Sutherland et al., 2006). Once we can assure that healthy deformation can be correctly

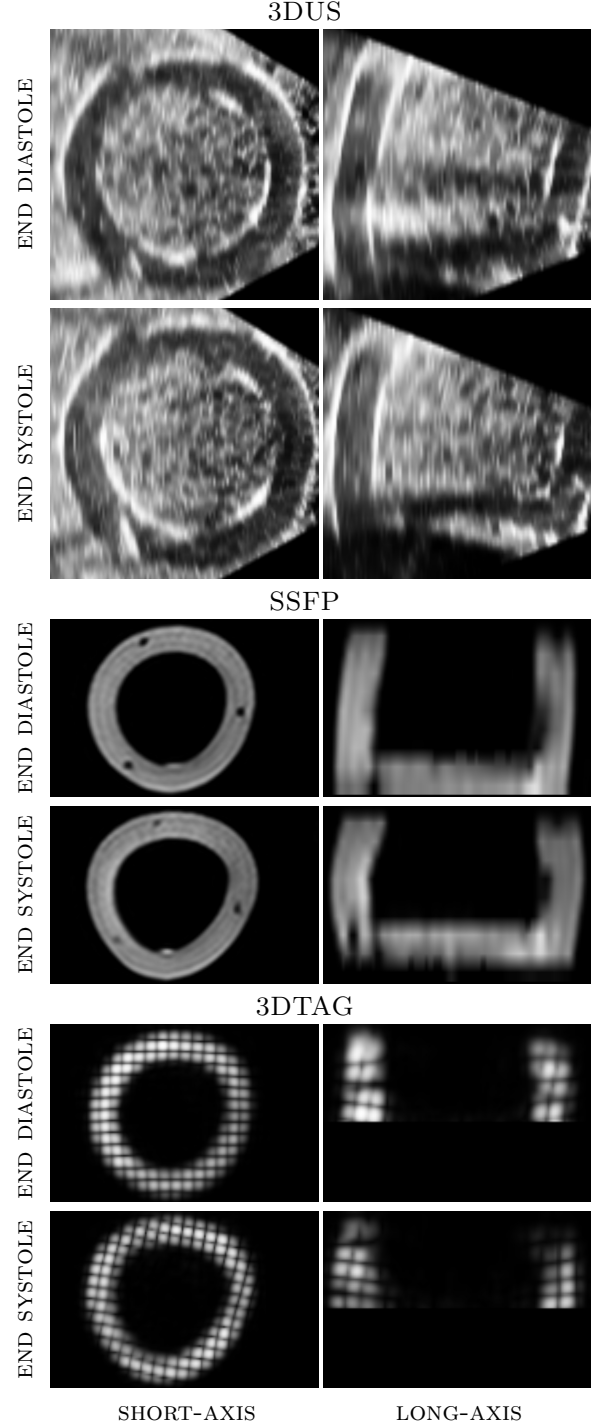


Figure 1: Dynamic phantom dataset. 3DUS= 3D ultrasound; SSFP= Steady State Free Precession; 3DTAG= 3D tagged Magnetic Resonance.

quantified by the algorithms, researchers can move further towards identification of pathological deformation from a new set of unseen data.

The collected database included 16 anatomies (15 volunteers + 1 phantom) in 2 modalities (see Sec. 2.2 and Sec. 2.3) and between 13-30 time frames. This amounted

⁸<http://www.cardiacatlas.org/>

Table 1: Dataset description: demographics, temporal and spatial resolution.

	Age	G	BSA	3DUS			sAX SSFP			3DTAG					
	yr	n.a.	m ²	pixel	thk	ph	sl	pixel	thk	ph	sl	pixel	thk	ph	
				mm	mm		mm	mm		mm	mm				
phantom	n.a.	n.a.		n.a.	1.35	0.96	19	9	1.20	8	30	70	1.01	1.01	23
V1	28	M	1.73		0.66	0.58	14	14	1.25	8	30	95	0.96	0.96	23
V2	30	F	1.55		0.66	0.58	16	11	1.25	8	30	80	0.96	0.96	29
V4	29	F	1.63		0.82	0.72	11	14	1.25	8	30	90	0.96	0.96	26
V5	36	M	1.84		0.77	0.68	15	14	1.25	8	30	94	0.96	0.96	23
V6	34	M	1.92		0.82	0.72	14	14	1.25	8	30	94	0.96	0.96	23
V7	32	M	1.99		0.82	0.72	17	14	1.15	8	30	100	0.96	0.96	31
V8	27	M	2.13		0.82	0.72	14	16	1.15	8	30	100	0.96	0.96	31
V9	29	M	1.78		0.82	0.72	14	14	1.25	8	30	94	0.96	0.96	30
V10	22	M	1.84		0.82	0.72	13	14	1.25	8	30	80	0.96	0.96	27
V11	22	M	1.88		0.82	0.72	15	14	1.15	8	30	100	0.96	0.96	32
V12	30	M	1.94		0.82	0.72	13	14	1.25	8	30	80	0.96	0.96	24
V13	31	M	1.78		0.77	0.58	24	14	1.15	8	30	90	0.96	0.96	38
V14	24	F	1.61		0.96	0.72	18	14	1.15	8	30	75	0.96	0.96	29
V15	20	M	1.65		0.96	0.72	13	12	1.25	8	30	90	0.96	0.96	21
V16	20	M	2.06		0.82	0.72	13	14	1.25	8	30	90	0.96	0.96	25

G= gender; BSA= body surface area; sAX= short-axis;
yr= years; pixel= in-plane pixel size; thk= slice thickness; ph= cardiac phases;
sl= slices; V3 = incomplete dataset.

to 1158 image volumes to be processed by each participant. Clearly, a considerable effort was required to participate in the challenge. Details on the database and acquisition protocols are provided below.

The phantom: the left ventricular phantom was constructed with polyvinyl alcohol (PVA, Lenticats, GeniaLab, Braunschweig, Germany). For MR acquisition, it was placed in an MR-compatible air-pressured actuator which compresses and rotates the PVA according to a preset heart rate (Manzke et al., 2010). For 3DUS acquisition, the phantom was inserted in a container filled with water to improve echogenicity conditions. The actuator was synchronized with the MR scanner and echocardiography system with an electrocardiography (ECG) waveform coupled to the standard ECG monitoring unit. Acquisition was performed at the Department of Internal Medicine II - Cardiology, University of Ulm, Germany.

The volunteers: sixteen healthy volunteers without clinical history of cardiac disease were recruited (3 female, aged 28 ± 5 years). The MR and 3DUS acquisitions were obtained within 3.5 ± 3.3 days of each other. One volunteer acquisition was incomplete, and therefore discarded (V3). Demographics and body surface area measurements can be found in Table 1. The data were acquired at the Division of Imaging Sciences and Biomedical Engineering, King’s College London, United Kingdom.

2.2. Ultrasound acquisition

The ultrasound datasets were acquired using an iE33 3D echocardiography system (Philips Healthcare, Best, The Netherlands) with a 3D X3-1 matrix array transducer. Full-volume acquisition (FVA) mode was used in which several smaller imaging sectors are combined to form a large composite volume. Each smaller sector was acquired in a single heart cycle. FVA was performed during breath-hold to minimize translation artifacts between the acquired sectors. Apical FVAs of the left ventricle were obtained in all subjects. The cardiologist was free to adjust

parameter settings to minimize background tissue signal and maximize frame rate (see Fig. 1 and Fig. 2). Details on spatial and temporal resolution of the datasets can be found in Table 1.

2.3. MR acquisition

The MR datasets were acquired using a 3T Philips Achieva System (Philips Healthcare, Best, The Netherlands). The MR sequences processed by the participants were cine Steady State Free Precession (SSFP) and 3D tagged Magnetic Resonance (3DTAG). SSFP datasets were scanned in multiple views (TR/TE= $2.9/1.5ms$, flip angle= 40°). All images were acquired during breath-holds of approximately 15 seconds and were gated to the vector ECG. 3DTAG datasets were obtained with three sequential breath-hold acquisitions in each orthogonal direction (TR/TE= $7.0/3.2ms$, flip angle= $19-25^\circ$, tag distance= $7mm$) (Rutz et al., 2008). Images were acquired with reduced field-of-view enclosing the left ventricle. A respiratory navigator was used to compensate for possible respiratory miss-alignment during the three sequential acquisitions (see Fig. 1 and Fig. 2). Details on spatial and temporal resolution of the datasets can be found in Table 1.

3. State of the art

Considerable efforts have been dedicated to quantification of cardiac motion from MR and US datasets. This section, without being an exhaustive survey of all published techniques, aims at situating the methods under comparison in a global context.

Three main categories are described for computing cardiac motion and strain. Indicative references for each category are given in Fig. 3. For an exhaustive list of references see Wang and Amini (2011); Petitjean et al. (2005) for cine and tagged MR modalities and to D’hooge et al. (2000); Angelini and Gerard (2006) for 3DUS.

Direct methods: the first category includes all methods that first detect features and further quantify motion and strain (see first block of Fig. 3). These methods were applied to 3DTAG images by detecting tag lines and tag intersections. Fiducial markers can also be used in other modalities (e.g. cine MRI) although the set of markers is likely to be too sparse for any detailed motion and strain analysis. 3DTAG images can be locally filtered to analyze the spatial frequencies and orientation of the tagging pattern. One example is the use of Gabor filters to derive strain directly. Although these methods avoid artifacts or excessive smoothing potentially induced by image registration, the computation of strain values per material points requires motion along the cycle to be quantified.

Fourier-based methods: the second category is based on the property that a local shift in the spatial domain corresponds to a phase shift in the Fourier domain. Several

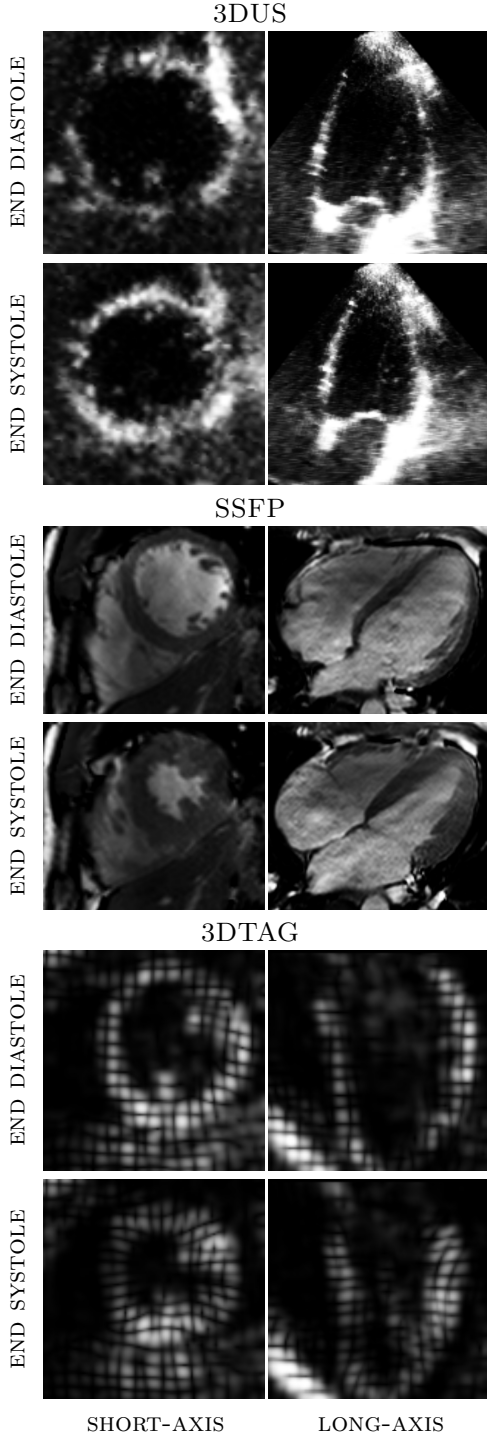


Figure 2: Average quality volunteer dataset. 3DUS= 3D ultrasound; SSFP= Steady State Free Precession; 3DTAG= 3D tagged Magnetic Resonance.

methods (see second block of Fig. 3) exploit this property by moving to the Fourier domain for locally quantifying the displacement field. Harmonic phases (HARP)-based techniques are a subset of this category and are restricted to the processing of tagged MR images. HARP models tagged MR images as the result of an amplitude mod-

ulation by a set of harmonics. For each harmonic, the complex phase of one material point is constant over the cardiac cycle. By doing band-pass filtering in the Fourier domain and coming back to the spatial domain, HARP images are computed. Displacement can then be calculated by division of space shift over phase gradient. The sinMod algorithm (Arts et al., 2010) is based on a similar concept as HARP. Next to computing a local phase shift, sinMod also computes local spatial frequency. The latter can be used for strain estimation in a similar manner as Gabor-based techniques.

Tracking-based methods: the third category (tracking-based approaches) solves for the displacement in the spatial domain by optimizing similarity between frame-to-frame intensities. Fig. 3 (third block) orders tracking-based methods following their transformation model. Spatial derivatives of the transformation yield strain estimates. In the case of Demons-based or block matching approaches, a spatial regularization is performed at each step to ensure smooth strain results. For other transformation models, transformation spatial derivatives can be obtained analytically.

The algorithms presented in this paper are highlighted in bold in Fig. 3. One algorithm (Tautz et al., 2012) is part of the Fourier-based category, while the remaining belong to set of tracking-based methods (Wang et al., 2012; De Craene et al., 2012b; McLeod et al., 2012). If we exclude techniques currently restricted to 2D images, our report misses direct methods such as Gabor filters. Since HARP is a reference method in tagged MRI processing, comparisons to 3D-derived HARP techniques would also add to this comparison. As both data and metrics of this paper are publicly available, we encourage future contributors to report results on our dataset.

4. Methodologies

The challenge raised interest from 13 worldwide groups: United States of America (4), Germany (2), The Netherlands (2), Canada (1), Egypt (1), France (1), Spain (1), United Kingdom (1). All of them successfully downloaded the data. After 3 official withdrawals, only 30% of the initially interested challengers submitted their results. Reasons for withdrawal included unreadiness of the methodology or inflexibility to input format.

We obtained 4 final submissions to the challenge from the following groups (by order of submission): Fraunhofer MEVIS (MEVIS), Bremen, Germany; Imperial College London - University College London (IUCL), UK; Universitat Pompeu Fabra (UPF), Barcelona, Spain; Inria-Asclepios project (INRIA), France.

A brief comparison of the four methodologies is shown in Table 2. Aspects related to the registration methodology employed by each challenger are summarized in Table 3. Two methods (INRIA and UPF) are extending classical registration methods by representing cardiac motion

Direct methods: detect first, then regularize.	
Tag beads (3DTAG, Amini et al. (2001); Kerwin and Prince (1998)) Gabor filters (3DTAG, Chen et al. (2010)) Fiducial markers (cine MRI, Remme et al. (2005))	
Fourier-based methods: measure phase shift in the Fourier domain.	
HARP, 3D-HARP and zHARP (2D and 3DTAG, see HARP references in Wang and Amini (2011)(Table 4)) SinMod (2DTAG, Arts et al. (2010)) Morphon (3DTAG and 3DUS, Tautz et al. (2012))	
Tracking-based methods: conservation of intensity.	
Block-matching	Sub-volume elastic (3DUS, Krucker et al. (2002)), block flow tracking (3DUS, Linguraru et al. (2007)), speckle tracking (3DUS, Crosby et al. (2009))
Demons-based	Optical flow (2DUS and 3DUS, see all references in Angelini and Gerard (2006); 3DTAG, Xu et al. (2010)), iLogDemons (3DTAG, 3DUS and cine MRI, Mansi et al. (2011) ; McLeod et al. (2012)), variational multimodal registration (3DTAG, Rougon et al. (2005); 3DTAG and cine MRI, Petitjean et al. (2004))
BSpline-based	3D BSpline (3DTAG, Deng and Denney (2004)), FFD (3DTAG, Chandrashekara et al. (2004a); 3DTAG and cine MRI, Shi et al. (2012) ; 3DUS, Elen et al. (2008))
BSpline-based (spatio-temp.)	4D BSpline (3DTAG, Huang et al. (1999)), 2D+t BSpline (2DUS, Ledesma-Carbayo et al. (2005)), 4D FFD (3DUS and 3DTAG, De Craene et al. (2012a,b)), 4D Nurbs (3DTAG, Tustison and Amini (2006))
FEM-based	Hyperelastic (cine MRI, Veress et al. (2002)), linear elastic (cine MRI, Papademetris et al. (2002))
RBF	Incompressible RBF (cine MRI, Bistoquet et al. (2008))

Figure 3: Classification of motion and strain quantification methods from cine MR, 3DTAG and 3DUS images. The methods evaluated in this paper are written in bold format.

by a velocity field. The temporal integration of velocity along the cardiac sequence gives the displacement, either frame to frame or from a reference frame to any other frame. By estimating the velocity field from the entire sequence, UPF’s algorithm introduces coupling between different time points. However, this comes at the cost of increased complexity. INRIA’s algorithm processes the sequence frame to reference with an added incompressibility constraint to enforce physiological transformations. Both approaches differ in the representation of the velocity field. Another difference between these two challengers is that INRIA opted for a dense field representation, whereas UPF represents the velocity field using a sparse set of B-Spline functions.

IUCL contributed an algorithm that fuses several MR sequences to recover a single estimate of the deformation. In this algorithm, displacement fields are represented by summing B-Spline functions. The main advantage of this method is to compensate for noisy or corrupted information in one sequence by adding complementary information from other MR protocols. Finally, MEVIS implemented a sparse method based on phase tracking and the Morphon

algorithm. As it processes the input sequence slice by slice (both in long and short axis), this approach is easy to parallelize. Phase-based methods are also conceptually close to HARP-based methods, being a prominent methodology in tagged image processing. Because of the periodic tagging patterns, Fourier-based method appear as a natural option for quantifying tag displacements.

In order to encourage participation, researchers were free to apply their methodology to as many modalities as possible. Either due to methodological or implementation constraints, some modalities were not processed by all challengers. In the remainder of this Section, we briefly describe each methodology along with their multimodality applicability.

4.1. MEVIS: Quadrature Filter Based Registration

The method used by Fraunhofer MEVIS is based on the Morphon algorithm introduced by Knutsson and Andersson (2005).

Background: the Morphon algorithm is based on the iterative quantification of displacement through phase shift estimate. It is based on the basic property that the Fourier

transform of a signal $f(x)$ and a shifted signal $f(x-d)$ are related via a phase factor by $\mathcal{F}\{f(x-d)\} = e^{-jd\omega} \mathcal{F}\{f(x)\}$. For two signals $f_1(x) = f(x)$ and $f_2(x) = f(x-d)$, d is proportional to $\arg\left(\mathcal{F}\{f_1(x)\}\overline{\mathcal{F}\{f_2(x)\}}\right)$, with $\overline{}$ denoting the complex conjugate. In practice, the phase shift can be estimated through a bank of quadrature filters. By combining the 1D displacement measured in the direction of each filter, the complete 3D displacement can be estimated. Another important feature of the algorithm is the estimation of a confidence measure, depending of the average response (both in amplitude and phase) of the intensity images to the filters. The total displacement estimate is obtained by averaging the shift from the different directions, weighted by the corresponding confidence weighting.

To estimate large deformations, the algorithm was implemented in a scale space, iterating the estimation several times on each scale. Confidence weight maps are accumulated over the iterations at each scale.

Implementation: to derive the motion fields from the 3DTAG datasets, the image data are analyzed slice-wise with a set of four 2D filters in orthogonal directions. Full 3D registration was not performed, unlike other challengers, for two reasons. First, the tagging grid is composed of tag lines in three orthogonal directions, and second, the computational costs for a 3D registration are significantly higher when compared to a 2D registration, making a distributed computation scheme highly desirable. Thus, the total motion field is obtained by combining three 2D motion vector fields from orthogonal directions, which can be computed independently. Because each 3D vector component is present in two fields, out-of-plane motion and other artifacts can be compensated for by combining the corresponding components.

The log-normal quadrature filters were applied in a six-scale scale space created by two different center frequencies and three resolution scale steps with a downsampling factor of 0.63. It should be noted that the frequencies are not related to the spatial frequency of the tagging grid, but were selected to give the scale-space scheme good coverage of motions with different magnitudes. To reduce computation time, the number of iterations was set to 4 at coarse scales and reduced to 2 at the finest resolution.

For each of the three orthogonal slice orientations, a stack of slice-wise 2-component vector fields is calculated. Each slice in this vector field stack represents the motion between two adjacent time points. These partial motion fields are reformatted to the transversal orientation. Every orientation contributes to two components of the final 3D vector that is determined by averaging the respective components. The 3DUS datasets are down-sampled by a factor of 2 for processing to compensate for their size and the increased noise level in comparison to the 3DTAG datasets, and up-sampled with linear interpolation afterwards.

Multimodal adaptation: the regularization of the deformation field was performed using a Gaussian kernel, whose parameters were empirically optimized. The Gaussian kernel is a function $G(x, \sigma) = e^{-\frac{x^2}{2\sigma^2}}$, where σ is 3 for all scales with down-sampled resolution, 4 for the 3DTAG original resolution, and 4.5 for the 3DUS original resolution. Because of the lower signal to noise ratios in 3DUS, the displacements were more strongly regularized in that modality.

4.2. IUCL: Multi-image Motion Tracking

Background: IUCL’s approach combines different MR sequences for the reconstruction of a single estimate of cardiac deformation. Such fusion strategy was motivated by observing that radial deformation can be clearly assessed from short-axis SSFP images. In 3DTAG images, radial motion is more difficult to quantify due to the limited number of tags, in comparison to longitudinal and circumferential directions. Therefore, combining different MR sequences in a common framework is expected to combine the advantages of SSFP and 3DTAG sequences.

Implementation: in IUCL’s implementation, myocardial motion was estimated from the first frame to each subsequent frame using a series of free-form deformations (Rueckert et al., 1999) as described in Chandrashekara et al. (2004b).

In a pre-processing step, all image sequences were resampled to a common temporal resolution using nearest neighbor interpolation. For spatial alignment, all SSFP sequences were brought in the 3DTAG coordinate system by rigid registration. Normalized mutual information was taken as similarity metric (Studholme et al., 1999) between the SSFP and the untagged 3DTAG intensities.

To balance the influence of the tagged and untagged sequences, a relative weight needs to be defined at each voxel. This weighting was designed to give more importance to 3DTAG intensities inside the myocardium and to SSFP intensities close to myocardial borders. The weight computation involved a gradient detector (applied to the SSFP image) and a probabilistic segmentation algorithm, returning at each voxel the probability to belong to myocardial tissue (Shi et al., 2011; Zhuang et al., 2010). The information from both segmentation and the gradient-detector were combined into a spatially varying weighting function moderating the influence of the tagged and untagged images on the motion tracking.

Cardiac motion was then estimated by mapping all images in the sequence to the first image. The transformation found at the previous frame was used as initial set of parameters in the optimization (Rueckert et al., 1999; Chandrashekara et al., 2004b).

Valve tracking was also integrated as a complementary source of information. The approach described in Zheng et al. (2008) was applied for joint detection and tracking of

landmarks from LA views. It first identifies regions which are likely to contain the landmarks of interest and then applies machine learning based landmark detector to voxels within this region.

Landmarks such as the valve annulus were tracked simultaneously in three LA views and a valve plane surface was constructed for each time frame using Delaunay triangulation. The distance to the constructed valve surface was applied as an additional penalty term in the registration cost function (Shi et al., 2012).

Multimodal adaptation: since IUCL’s approach is tailored to fuse information from various MR sequences, it was difficult to extend it to ultrasound. Possible extensions of this algorithm to ultrasound, as well as related algorithmic modifications, are described in the Discussion.

4.3. UPF: Temporal Diffeomorphic Free Form Deformation (TDFFD)

The TDFFD algorithm described in De Craene et al. (2012a) optimizes a 4D velocity field parameterized by B-Spline spatiotemporal kernels. The advantage of representing the velocity rather than the displacement is to introduce temporal consistency in the recovered transformation. That is, that motion at a given time point depends on all previous times.

Background: the original metric in the TDFFD algorithm was summing squared intensity differences between the first frame and each subsequent frame in the sequence. In this paper, an additional term measuring frame to frame differences was added to the metric. This term is expected to be more sensitive to small incremental displacements. It is also more robust to possible intensity changes due to smooth alterations of tissue magnetization in time. However, it can generate a significant drift effect by accumulating small tracking errors over the whole cycle. Therefore, sequential and *absolute* (i.e. to the reference) intensity differences need to be combined. In the TDFFD algorithm, modifying velocity at one time point impacts all intensity differences between the next frames and the reference frame. This tends to add *coupling* between frames.

Implementation: the implementation used in this paper is publicly available from De Craene and Piella (2012). Samples are taken randomly in a 3D mask specified over the entire sequence. The mask was adjusted to contain myocardium over the entire sequence some sources of artifacts (e.g. diaphragm in the 3DTAG sequences). The design is flexible for plugging different similarity metrics and adjusting the weights between the different terms. In this paper; the balance between the two terms was chosen to have at initialization an equal contribution of the two terms.

Multimodal adaptation: on the one hand, to process 3DTAG datasets, the tracking is performed backwards in

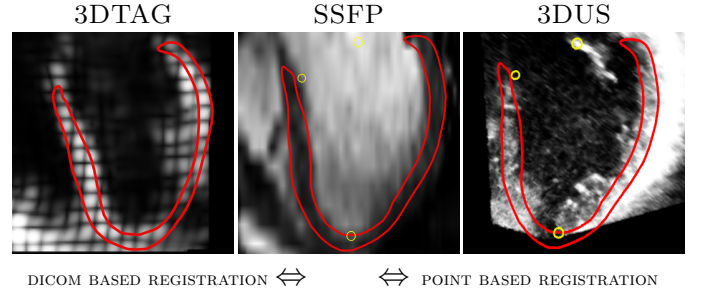


Figure 4: A segmentation is obtained from the short-axis SSFP dataset (middle). The LV model is registered to 3DTAG coordinates using DICOM header information (left) and to 3DUS coordinates using a point based similarity transform (right). Red= contours of segmented LV; Yellow= landmarks used for point based registration.

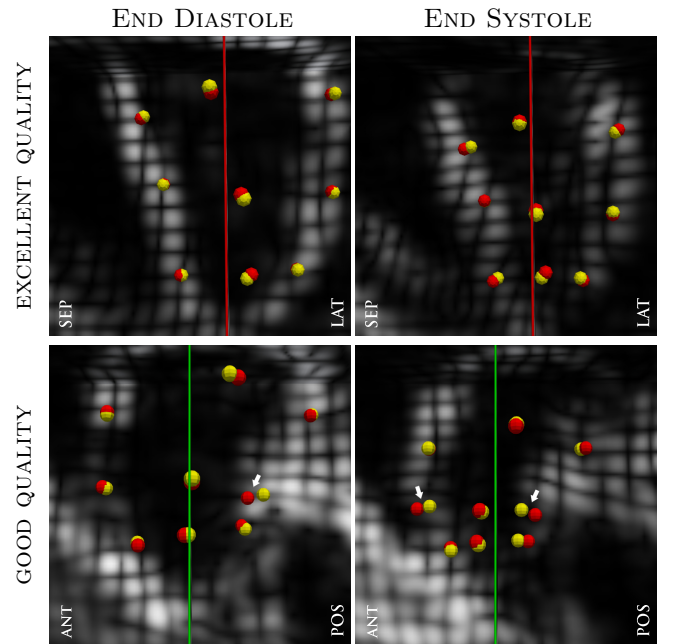


Figure 5: Results of manual landmark tracking in two datasets: one with good and one with excellent quality. Note the in-out-of-plane motion of the landmarks due to our 4D tracking strategy. Different colors represent different observers. The white arrows show regions with high inter-observer variability (see animation in supplementary material). Yellow= landmarks tracked by observer 1; Red= landmarks tracked by observer 2. Sep= septal wall; Lat= lateral wall; Ant= anterior wall; Pos= posterior wall.

time starting from the last image. This option was taken because in the first images of a 3DTAG sequence, the blood is magnetized in the same way as the tissue. Performing tracking from the last frame was therefore a better choice of reference given the lack of tags in the ventricular cavities. On the other hand, 3DUS and SSFP datasets do not present this complication. Therefore, the first frame was taken as reference and tracking was done forward.

4.4. INRIA: Incompressible Log-Domain Demons

The incompressible log-domain demons algorithm described in Mansi et al. (2011) (iLogDemons for short) is an optical-flow method for co-registering images that was adapted from the previous log-domain Demons algorithm (Vercauteren et al., 2008) for cardiac tissue tracking to impose physiological constraints in the myocardium such as incompressibility and elasticity.

Background: the iLogDemons algorithm is an extension of the LogDemons algorithm proposed by Vercauteren et al. (2008). As in the original LogDemons, the transformation is parameterized by stationary velocity fields through the exponential map (Arsigny et al., 2006). Squared intensity differences between two frames are minimized iteratively. Each iteration includes an optimization step, solving for the gradient of the image-term, and a regularization step. The latter is performed by an elastic-like regularizer involving the Hessian of the Gaussian kernel. Adding incompressibility of the myocardium is enforced by constraining the velocity fields to be divergence-free. This projection operation is performed by adding incompressibility as a Lagrange multiplier, minimizing the norm of the difference to the original velocity field.

Implementation: in order to optimize the results of the tracking, some image pre-processing was applied. First, the images were resampled to give isotropic voxel sizes. Image contrast was enhanced by trimming the tails of the grey-level histogram to exclude the 1st and 99th percentiles. In order to impose incompressibility in the myocardium, a mask of the myocardium at the reference frame (end diastole) was required. A full sequence was registered by computing the deformation from each frame to the reference frame. The transformation was initialized by the result of the previous frame. The use of a fixed reference frame allows to define the mask of the myocardium once for the whole sequence. For more details on the sequence tracking pipeline, see Mansi et al. (2011).

Multimodal adaptation: the algorithm was applied identically to all imaging modalities. The only difference was related to the mask definition for imposing incompressibility. In the case of the 3DUS and 3DTAG image sequences, only the mask of the left ventricle myocardium was used in the registration since the right ventricle was not clearly visible in all cases for these modalities. For the SSFP and 3DUS sequences, the first frame was chosen as the reference frame, since the first frame is assumed to correspond to end diastole. For the 3DTAG sequences, the third frame was used as the reference frame as this was the first frame in the sequence with the blood pool distinguishable from the tissue.

5. Evaluation

5.1. Left ventricular mesh

In order to visualize the deformation fields provided by the participants, we obtained anatomical information from the SSFP datasets as follows: 1) we selected the short-axis SSFP frame with closest trigger time to the end diastolic 3DTAG frame; 2) we segmented the left ventricle (LV) from the selected short-axis dataset by manually deforming a left ventricular model; 3) we registered the segmented LV mesh to 3DTAG coordinates using DICOM header information (see Fig. 4).

The LV model extracted from the SSFP dataset was registered to 3DUS coordinates as follows: 1) we selected three orthogonal visualization planes to match typical MR acquisition planes; 2) we marked three anatomical landmarks on the four-chamber view; 3) with the corresponding landmarks on the MR datasets, we performed a point based similarity transform on the LV mesh (see Fig. 4).

5.2. Ground-truth landmarks

In order to generate ground truth landmarks from 3DTAG datasets, we selected 8 landmarks for the phantom and 12 landmarks per volunteer: one landmark per wall (anterior, lateral, posterior, septal) per ventricular level (basal, midventricular, apical). Each landmark was placed on the intersection of the tag lines on the three orthogonal visualization planes. These landmarks were used as initialization points and manually tracked by two observers. In order to achieve real 4D tracking of the landmarks, the process was done one landmark at a time. That is, the following time frame was visualized and the landmark was manually displaced to follow the tag marks on the three orthogonal planes. This process was repeated until the landmark was tracked over the whole cardiac cycle (see Fig. 5 and animation in supplementary material). Visualization, segmentation and tracking were performed with GIMIAS v1.4.0 (Larrabide et al., 2009). Conversion to SSFP coordinates was done with DICOM header information. For 3DUS, we mapped the landmarks from the 3DTAG datasets to 3DUS coordinates using the mapping described in Sec. 5.1. The manually tracked landmarks were used to evaluate tracking accuracy of all methodologies. Statistical significance of the medians was assessed with a Mann-Whitney test.

5.3. Strain computation

The myocardial strain tensor was directly estimated from the spatial derivatives of the displacement field φ_0^n , mapping the reference frame to a frame n in the sequence. To apply the same computation technique for dense field and B-Spline based approaches, we generated the sequential displacement fields for each method and composed them for computing the set of φ_0^n corresponding to each method and modality. Strain was then computed as

$$\epsilon(\mathbf{x}, t_n) = \frac{1}{2} \left((\mathbf{D}\varphi_0^n)^T \mathbf{D}\varphi_0^n - \mathbf{I} \right) \quad (1)$$

Table 2: Features of each methodology (excluding tracking algorithm)

INPUT	MODALITY	PREPROCESSING		PROCESSING		SPECIFICITIES		
		DESCRIPTION	TIME P.C.	DESKTOP	CLUSTER	STRENGTHS	WEAKNESSES	
MEVIS	VTK URG	3DTAG	n.a.	n.a.	1-2 hrs	n.a.	Purely data-driven; Can be parallelized (3 orientations on 2D slices);	Difficult to include prior information (e.g. tissue elasticity);
		3DUS					Independent of image dimension, modality and target anatomy.	
IUCL	DCM	combined 3DTAG SSFP	auto segm.	20 min	3-6 hrs	2-4 hrs	Complementary information; Robust against accumulative error; Automatic segm.	Spatial misalignment; No incomp. constraint; Smooth constraint avoids large deformation.
								High computation time;
UPF	VTK URG	3DTAG	mask	10 min	6 hrs	6 hrs	Time consistency;	Temporal smoothing difficult to modulate with current implementation.
		3DUS SSFP					Robustness to low SNR.	
INRIA	ITK Image*		mask	20 min	n.a.	5 hrs	Controlled incomp;	Incomp. fully constrained; Requires seg; Regularization strength difficult to assess per patient.
							Suitable for all modalities;	
							Physiological priors;	
		3DTAG					Volume variation <5%;	
		3DUS					Computationally efficient (GPU).	
		SSFP						
p.c.=per case; n.a.= not applicable; URG= uniform rectilinear grid; *= generated from VTK files; Segm= segmentation; Incomp= incompressibility.								

p.c.=per case; *n.a.*= not applicable; URG= uniform rectilinear grid; * = generated from VTK files; Segm= segmentation; Incomp= incompressibility.

Table 3: Tracking features of each methodology

	PRINCIPLE	DOMAIN	PARAMETRI- ZATION	DISP./VEL.	REGULARIZATION	IMAGE FORCE	SEQUENTIAL VS. ABSOLUTE REFERENCE [†]		
							DIRECTION	METRIC	TRANSFORM
MEVIS	Quadrature filter-based elastic registration	3D (Sparse)	Dense	Disp.	Confidence maps	Phase difference	Forward	Sequential	Sequential
IUCL	Cross Correlation based Free Form registration	3D	FFD	Disp.	Bending energy	Mutual information	Forward	Absolute ref.	Absolute ref.
UPF	Temporal Diffeomorphic Free Form deformation	4D	FFD	Vel.	Coupling from adjacent time points	Intensity difference	Backward (tagged MR) and Forward (cineMR/3DUS)	Sequential	Sequential
INRIA	iLog Demons	3D	Dense	Vel.	Incompressibility	Intensity difference	Backward	Seq. then abs.	Seq. then abs.

[†]= backward if time and transformation directions are opposite; Disp.=displacement; Vel.=velocity.

where the \mathbf{D} operator computes the spatial derivatives of the φ_0^n transformation by finite differences. For computing Lagrangian strain, the displacement field maps points from the ED frame to each frame of the sequence. Strain was projected along a specific direction of a local coordinate system related to the anatomy of the LV. The local coordinate system was computed using $\epsilon_p(\mathbf{x}, t_n) = \mathbf{p}^T \cdot \epsilon(\mathbf{x}, t_n) \cdot \mathbf{p}$ where p represents the radial, longitudinal, circumferential direction. The longitudinal direction, \mathbf{l} , was defined uniformly by drawing a line from the apex to the mitral valve. These two landmarks were consistently stored on the mesh, making long axis computation automatic and robust to any image orientation. The radial direction was obtained from the normal \mathbf{e} to the mesh at each node and the \mathbf{l} vector using $\mathbf{r} = \mathbf{e} - (\mathbf{e} \cdot \mathbf{l})\mathbf{l}$. The circumferential direction was then obtained by the cross product of longitudinal and radial directions.

Strain values are usually visualized for each of the 17 American Heart Association (AHA) segments. However, the definition of the segments on each patient-specific geometry is a challenge in itself. To tackle this issue we use the approach presented in De Craene et al. (2012b). This approach provides a regular definition of the AHA segments by remeshing the LV surfaces as follows: 1) we extracted the endocardial surface and mapped it onto a disk; 2) we corrected the mapping by moving the apical point to the center; 3) we created a new parametrization that maps the new vertices onto the surface; 5) we added the regions to the surface from a disk definition of the AHA segments; 6) we transferred each AHA label to

its corresponding closest epicardial vertex (for further details see De Craene et al. (2012b)). Finally, to visualize strain curves for each anatomical direction, we averaged strain values over several AHA segments and plotted them as a function of time. Aiming at highlighting differences between methodologies, the strain curves were not post-processed, smoothed or compensated for drifting errors. Results are presented in Sec. 6.4 and further discussed in Sec. 7.3.

5.4. Quality assessment

In order to study the influence of image quality on the methodologies, we have obtained quality gradings for both MR and 3DUS datasets. These gradings were used to subgroup the datasets and analyze accuracy tendencies.

MRI: each MR sequence was assessed by consensus of two expert cardiac MR researchers: a junior one (5 years experience) and a senior one (10 years experience). All scans were of clinically diagnostic image quality with full coverage of the left ventricle going from the apex to just beyond the mitral and aortic valve planes. Scores were given on a scale of 1 (average) to 3 (excellent). The scores were given as follows: 1= localized artifacts in more than one slice or non-localized artifacts in one slice; 2= localized artifacts in no more than one slice; 3= no artifacts in any of the slices throughout the left ventricle.

3DUS: datasets were assessed by consensus of two expert cardiologists four weeks after the last recruitment.

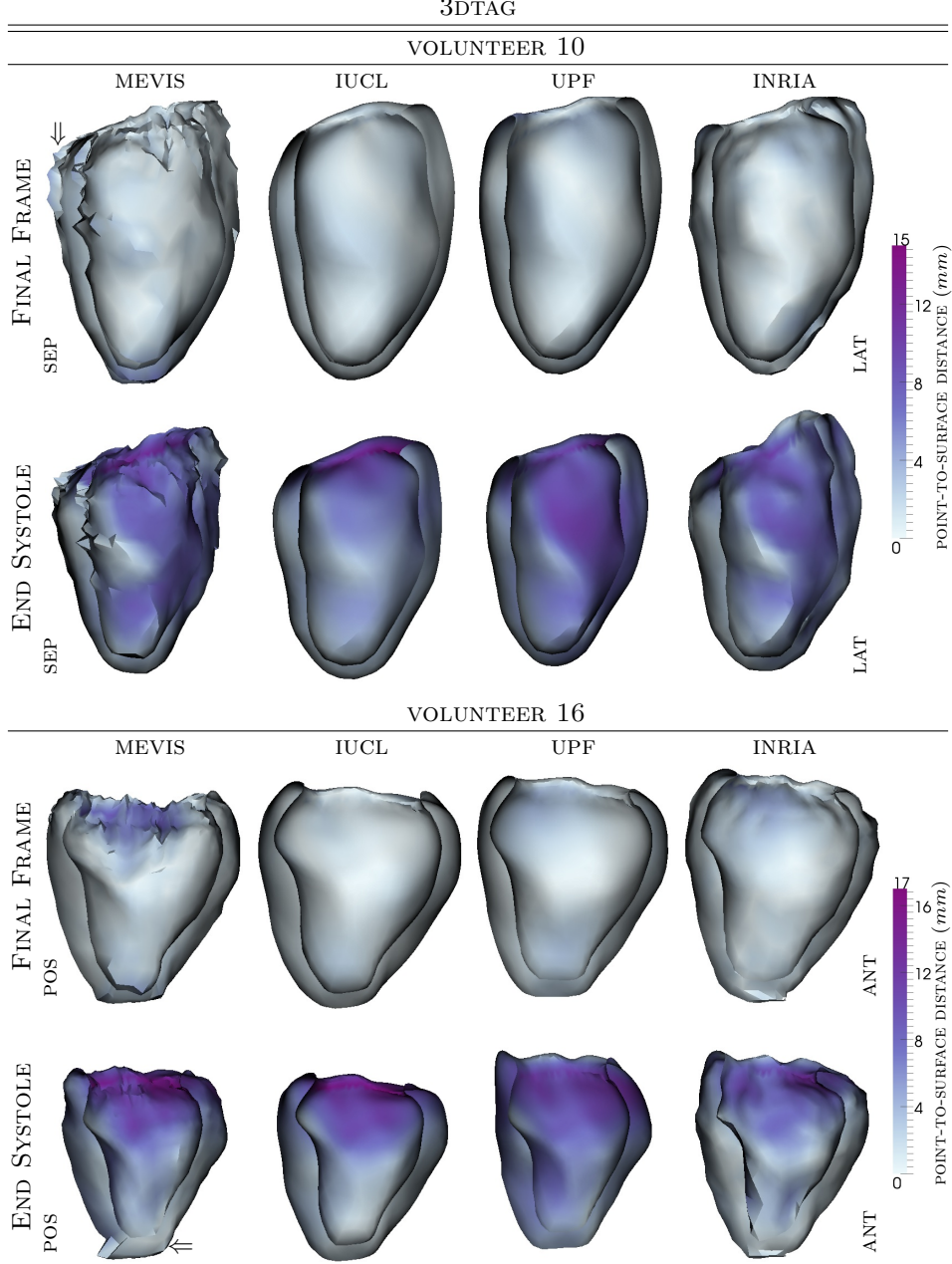


Figure 6: Deformed meshes for each challenger. The meshes are color mapped with point-to-surface distances to the reference mesh (manually segmented). Results are shown for the end systolic frame (bottom) and the final frame (top): last frame for forward tracking methods, and, first frame for backward tracking methods (see animation in supplementary material). Sep= septal wall; Lat= lateral wall; Ant= anterior wall; Pos= posterior wall.

To avoid bias, datasets were renamed with a randomized ID. The assessment was performed at global (overall score) and regional level (each myocardial AHA segment). The scores were given as follows: 0= unusable dataset; 1= usable dataset but with difficulty/reduced certainty of results; 2= average quality/usable dataset; 3= good dataset.

6. Results

In this section we present our resulting figures and plots. They are further discussed in Sec. 7.

6.1. Inter-observer variability

After obtaining the ground-truth as described in Sec. 5, we calculated the inter-observer variability. The obtained inter-observer errors were analyzed under two criteria. *Criterion 1*: the final position of the landmark was *relatively*

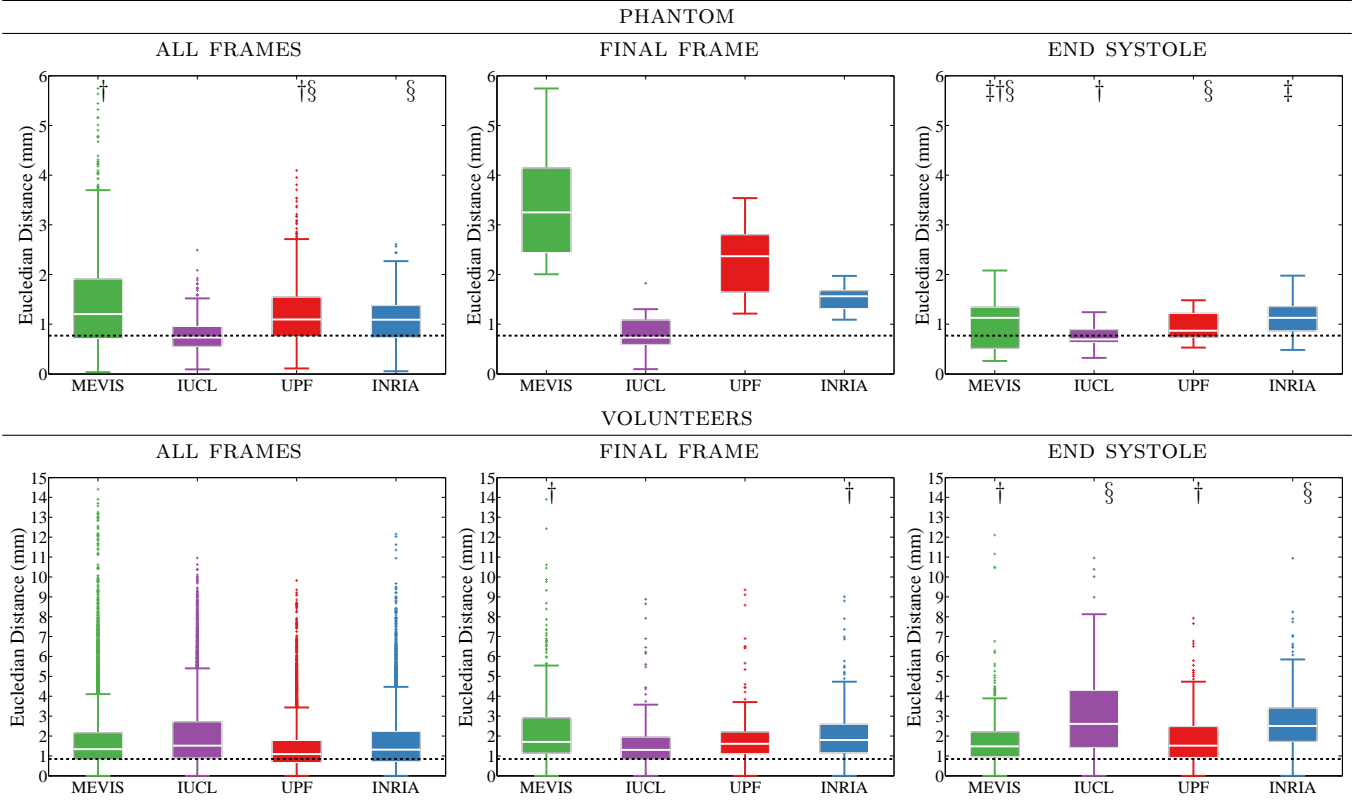


Figure 7: Box-plots of tracking errors for each challenger for **3DTAG** datasets of the phantom (top) and the volunteers (bottom). Maximum whisker corresponds to approximately 99.3% coverage if the data were normally distributed. Dotted black line represents average inter-observer variability. Pair of samples that yielded statistically non-significant differences ($p < 0.05$) are marked on the plot (†, ‡, §).

close to the initial position. We can assume the latter since all datasets are from healthy volunteers who are expected to have cyclic motion. *Criterion 2*: the final positions suggested by the two observers were *relatively close*. A landmark was labeled as *relatively close* when the distance was below the 75th percentile of all measured distances. Landmarks that did not follow both criteria were excluded from further quantification. The median of the inter-observer variability (IV) was computed over all time frames for the phantom dataset (0.77mm) and for the volunteer datasets (0.84mm).

6.2. Qualitative Analysis

To visualize the results obtained by each challenger, we applied the deformation fields to the manually segmented meshes. The obtained results can be found in Fig. 6 (see additional material for an animated version).

6.3. Tracking accuracy

3DTAG: median values were computed over all time frames for the phantom dataset (MEVIS=1.20mm, IUCL=0.73mm, UPF=1.10mm, INRIA=1.09mm) and for the volunteer datasets (MEVIS=1.33mm, IUCL=1.52mm, UPF=1.09mm, INRIA=1.32mm). Box-plots of tracking

accuracy errors are displayed in Fig. 7. This figure also includes accuracy plots for end systolic (ES) frame and final frame (FF). The final frame corresponds to: last frame for forward tracking methods, and, first frame for backward tracking methods. As supplementary material, we have provided animations of landmark tracking for all participants. Fig. 9 shows tracking accuracy errors per each subgroup as suggested by the quality assessment evaluation.

3DUS: due to temporal miss alignment between 3DTAG and 3DUS, accuracy errors were only evaluated at FF and ES. Median values of these two phases were computed for the phantom dataset (MEVIS=4.40mm, UPF=3.48mm, INRIA=4.78mm) and for the volunteer datasets (MEVIS=3.51mm, UPF=3.71mm, INRIA=4.07mm). Box-plots of tracking accuracy are displayed in Fig. 8. Fig. 9 shows tracking accuracy errors per each subgroup as suggested by the quality assessment evaluation.

SSFP: similarly to 3DUS, due to temporal miss alignment between 3DTAG and SSFP, accuracy errors were only evaluated at FF and ES. Median values of these two phases were computed for the phantom dataset

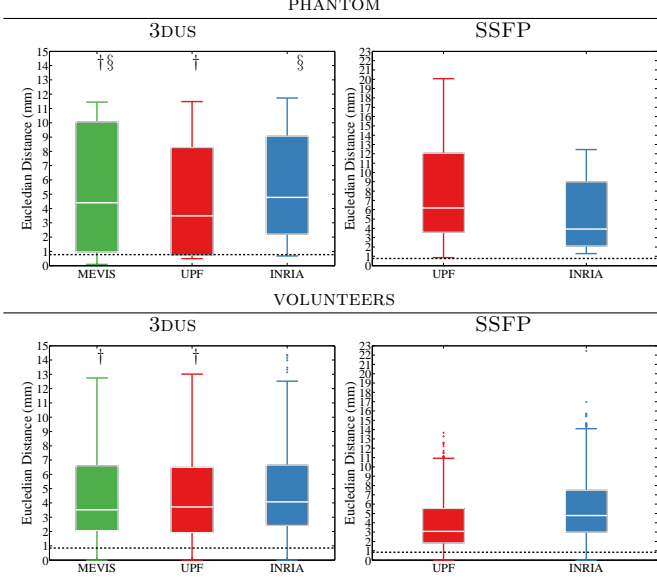


Figure 8: Box-plots of tracking errors for each challenger for **3DUS** and **SSFP** datasets of the phantom (top) and the volunteers (bottom). Due to temporal miss-alignment accuracies were only computed for FF and ES. Maximum whisker corresponds to approximately 99.3% coverage if the data were normally distributed. Dotted black line represents average inter-observer variability on the 3DTAG data. Pair of samples that yielded statistically non-significant differences ($p < 0.05$) are marked on the plot (†, ‡, §).

(UPF=6.18mm, INRIA=3.93mm) and for the volunteer datasets (UPF=3.09mm, INRIA=4.78mm). Box-plots of tracking accuracy errors are displayed in Fig. 8. Fig. 9 shows tracking accuracy errors per each subgroup as suggested by the quality assessment evaluation.

6.4. Strain curves

Strain measurements obtained for V9 are displayed in Fig. 10. This figure displays each strain component color mapped on the deformed mesh, corresponding to each challenger. The colormap diverges from a neutral color (bone) representing zero values, to blue colors representing negative values (as expected from longitudinal and circumferential components) and red colors representing positive values (as is expected from the radial component).

Plotting strain curves for each methodology and for each AHA segment would generate a large amount of figures to present in this manuscript. To give a concise view of strain quantification results, we opted for averaging strain curves in two regions: along the septum and along the lateral wall. The septum is usually placed in the center of the acquisition ROI and lacks surrounding structures that might create artifacts. As a result, it is consistently of higher quality than other myocardial walls. On the other hand, the lateral wall usually presents artifacts, such as out-of-ROI effect and surrounding structures interference. Fig. 11 displays strain curves obtained for volunteers V4, V9, V13 and V16 from all the modalities. To give an idea of image quality, a vignette of each

dataset at end systole is shown at the upper-left corner of the strain plots. Keep in mind that not all modalities were processed by all challengers. Therefore, the plots present: 4 curves for 3DTAG (MEVIS, IUCL, UPF, INRIA), 3 plots for 3DUS (MEVIS, UPF, INRIA), and, 2 plots for SSFP (UPF, INRIA). These cases were selected since they highlight strengths and weaknesses of all algorithms. It should be noted that strain curves were not post-processed, smoothed or compensated for drifting errors. Any noticeable smoothness is given by the methodology itself.

7. Discussion

7.1. Qualitative Analysis

Visual rendering of the results can reveal some advantages and drawbacks for each method. Since the MEVIS method works in a purely sequential manner, errors are accumulated over the cycle and induce irregularities (drift effects) on the deformed mesh at the end of the cycle. Another noticeable effect is the lack of deformation in areas that expand beyond the field-of-view. This happens often on the lateral wall and the apex (see Fig. 6-bottom). On the other hand, because both IUCL and UPF methods use a fixed reference in the similarity metric, they show good anti-drift capabilities. This is evident on the resemblance of the ED mesh to the reference mesh (small point-to-surface distances). Both IUCL and INRIA methodologies obtain large radial deformations. For IUCL, it is due to the inclusion of SSFP data on the pipeline. For INRIA, it is due to the incompressibility constraints, which was imposed for this purpose. This suggests that information derived from 3DTAG images alone is insufficient for accurately quantifying radial myocardial deformation. Most likely, due to coarse resolution of the tag spacing, radial deformation will more easily be seen on the short-axis SSFP dataset than on the 3DTAG dataset.

7.2. Tracking accuracy

3DTAG: due to DICOM miss-alignment of the phantom SSFP and 3DTAG datasets (Fig. 7-top), IUCL results include only 3DTAG information. In this case their accuracy error was lower than for the remaining methodologies. Both MEVIS and UPF obtained very different accuracy errors at FF and at ES. This indicates that their lower performance on this dataset arises from tracking errors at the end of the cycle. This difference is not evident on the volunteer datasets (Fig. 7-bottom). Most likely due to the high symmetry of the phantom dataset, the tracking becomes more challenging for MEVIS and UPF methodologies. Volunteer datasets, on the other hand, present enough anatomical landmarks able to penalize drift in the similarity metric. IUCL and INRIA were more robust to this effect in the phantom dataset. For the volunteers, INRIA obtained similar errors both in FF and ES. IUCL, on the other hand, obtained noticeably larger tracking errors

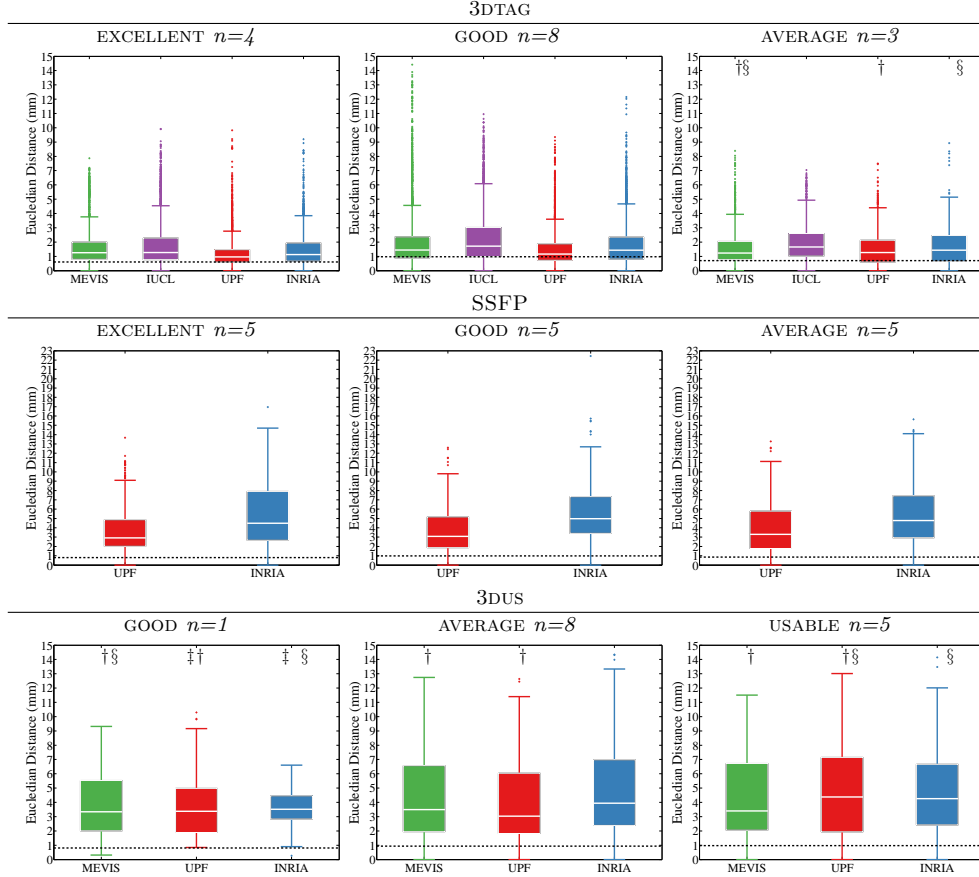


Figure 9: Box-plots of tracking errors for each challenger for 3DTAG (top), SSFP (middle) and 3DUS (bottom) datasets grouped by **quality grading**. One of the 3DUS dataset graded as unusable was discarded. Maximum whisker corresponds to approximately 99.3% coverage if the data were normally distributed. The dotted black line represents average inter-observer variability. Pair of samples that yielded statistically non-significant differences ($p < 0.05$) are marked on the plot (†, ‡, §).

at ES. As for the subgroup analysis, we mainly observe slightly smaller median errors on the excellent datasets for all methodologies.

3DUS: for the phantom, UPF results were significantly lower than those of INRIA. MEVIS results were not significantly different than UPF and INRIA. In the phantom, the symmetry of the cylindrical shape combined with low signal and speckle make the tracking specially challenging. For the volunteers, MEVIS results obtained a noticeable reduction of outliers (w.r.t. 3DTAG). This is most likely due to stronger constraints imposed on this modality to cope with lower image quality. On the other hand, INRIA results obtained higher number of surface irregularities, evident on the amount of outliers on the box-plot. This could indicate a need for increasing the smoothing parameters on the methodology. Average errors were lower for MEVIS and UPF. For this modality, dataset quality has a stronger influence on algorithm performance (Fig. 9). For an excellent dataset ($n = 1$), all methodologies obtained comparable results. For average datasets ($n = 8$), both MEVIS and UPF obtained lower accuracy errors than INRIA. For usable datasets ($n = 5$), UPF and INRIA ob-

tained similar errors, while MEVIS obtained slightly lower errors. Besides overall image quality, 3DUS suffers from reverberation artifacts inducing signal drops in part of the image domain. Another challenge in 3DUS is the presence of a bright pericardium next to a darker myocardial tissue in the lateral wall. Since the pericardium does not move at the same speed as the myocardium, this discontinuity tends to disturb registration algorithms.

SSFP: box-plots of tracking accuracy errors are displayed in Fig. 8. For the phantom, INRIA obtained significantly lower errors than UPF. This is most likely due to the high symmetry found in the SSFP phantom datasets (see Fig. 1). In this case the constraints present in INRIA methodology yield higher tracking accuracy. For the volunteers, on the other hand, UPF obtained significantly lower errors than INRIA. As for the subgroup analysis, we observe slightly smaller median errors on the excellent datasets for both methodologies.

7.3. Strain curves

Analyzing Fig. 10, we can clearly observe the smoothing effect of BSpline-based methodologies (IUCL and UPF).

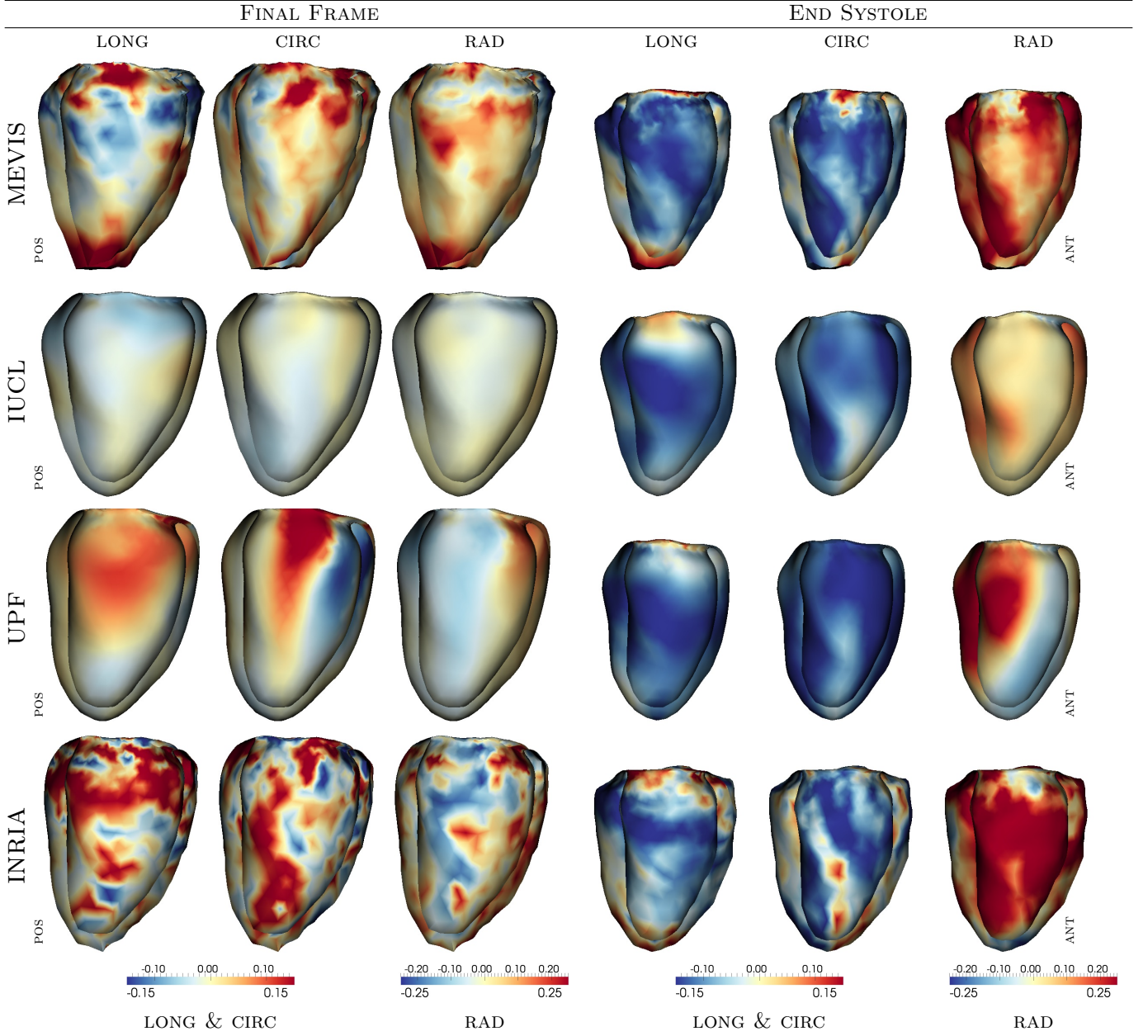


Figure 10: Strain measurements computed from the 3DTAG dataset of V9. Each strain component (columns) is visualized on the lateral wall for each participant (rows). The colormap diverges from a neutral color (bone) representing zero values, to blue colors representing negative values (as expected from longitudinal and circumferential components) and red colors representing positive values (as is expected from the radial component). Long= longitudinal strain, Circ= circumferential strain; Rad= radial strain; Ant= anterior wall; Pos= posterior wall.

This smoothness is paid off by decreased values of radial strain. Both for MEVIS and INRIA we observe more irregularities on the meshes. These irregularities are more evident at FF, due to accumulation of errors. However, the peak values are consistent across methods. Moreover, the calculation of radial strain is closer to physiological values (45%) for these two participants (MEVIS and INRIA). This suggests that the lack of regularization allows

the methods to track larger deformations (i.e. in the radial direction). If desired, these methods can be regularized to obtain smoother results. Regarding the different strain components, we can observe two main behaviors. On the one hand, longitudinal and circumferential ES strain maps were more uniform across methodologies. On the other hand, radial strain showed high variability at ES, with underestimated peak values by BSpline-based methodologies

(IUCL and UPF). This is further discussed below taking into account strain curves from different modalities.

Analyzing Fig. 11 for each modality, we observe the following. For 3DTAG, forward tracking techniques (MEVIS, IUCL) display the drift effect at the end of the cycle (Fig. 11-left). For MEVIS, the effect is more noticeable on the lateral wall (V4 and V13). For V4, we can see in the 3DTAG vignette that this is due to out-of-ROI effect. For V13, we can see in the 3DTAG vignette that this is due to image artifacts (most likely breathing). For backward tracking techniques (UPF, INRIA), the drift effect is noticeable at the beginning of the cycle (V4-septum, V16-lateral). Apart from these punctual drift effects, there is good agreement between participants on circumferential and longitudinal strain curves. The main differences can be observed on radial strain curves. Due to the incompressibility constraints, INRIA obtained radial strain peak values more consistent with those reported in literature (Moore et al., 2000). Due to the inclusion of SSFP data, IUCL also obtained positive curves (for most cases). MEVIS and UPF are more dependent on the dataset for the computation of radial strain. Reiterating, this may indicate that current 3DTAG protocols might be insufficient for accurately quantifying radial strain.

For 3DUS datasets, the overall shape of longitudinal and circumferential strain are retrievable only on datasets with good quality. However, the temporal resolution is crucial in strain calculation. On the one hand, we can see how in the dataset with maximum temporal resolution (V13, 24 cardiac phases) the strain curves are very similar to the ones obtained from 3DTAG and SSFP. On the other hand, in a dataset with good quality but lower temporal resolution (V16, 13 cardiac phases) the obtained curves are over-smoothed (see 3DUS-vignette). For radial strain computation, we observe the same differences as in 3DTAG. For this modality, the drift effect of MEVIS is less noticeable, while peak values are lower. This suggests a stronger smoothing constrain during computation.

For SSFP datasets, longitudinal and circumferential strain curves are best captured by UPF. Radial strain curves are again best captured by INRIA (Fig. 11-right). Note that in cases where the longitudinal and/or circumferential strain are underestimated (V16-lateral), due to a fully constrained incompressibility, the radial strain is overestimated. For some cases, the radial strain curves are comparable for both approaches. This suggests that with this modality radial deformation is indeed retrievable.

Finally, according to reported measurements (Moore et al., 2000), peak values of strain curves should reach up to 45% for radial strain, -20% for circumferential strain and -16% for longitudinal strain. Hence, peak values measured in this study, seem to be underestimated for all methodologies. However, the time and duration of the peaks are consistent per each volunteer across modalities (except in low quality 3DUS datasets). Radial strain is only retrievable from SSFP data. In this sense, the inclusion of incompressibility constraints are beneficial to obtain a physiologically

plausible strain curve.

7.4. Limitations of the study

Holding a challenge brings about the risk of misrepresenting the state of the art. The amount of reported methodologies should be viewed in the context of the level of difficulty of the challenge. Firstly, our challenge included 2 imaging modalities (i.e. MRI and US). Secondly, it included a variety of imaging protocols, from standard (i.e. multislice SSFP) to novel acquisition protocols (i.e. 3DTAG and 3DUS). The level of difficulty is therefore quite considerable. Nonetheless, we present 4 methodologies that have been able to tackle this level of difficulty, covering Fourier-based, Demons-based and BSpline-based methods (see Sec. 3 for methods categorization). In order to encourage future evaluations, all data, ground-truth and evaluation metrics are publicly available for download via the Cardiac Atlas Project.⁹

Regarding the realism of the phantom data, while PVA is ideal for MR applications, it features ultrasound attenuation coefficients below ~ 0.3 dB. This results in echogenicity signal which is lower than tissue, expected to have attenuation coefficients around ~ 1.0 dB (Surry et al., 2004). Nonetheless, the attenuation difference between the phantom and the surrounding water makes the edges of the phantom clearly visible. Therefore, while all methodologies were able to track the borders, torsion inside the volume was not retrievable due to low speckle signal.

8. Conclusion

This manuscript presents the results of the *Cardiac Motion Analysis Challenge* held at the STACOM 2011 workshop. The challenge consisted of quantifying myocardial motion from a phantom dataset and 15 volunteer datasets. Each set included 3DTAG, cine SSFP MRI and 3DUS images. Computed results were evaluated against manually tracked landmarks. Four institutions responded to the challenge by providing motion estimates for the data: Fraunhofer MEVIS (MEVIS), Bremen, Germany; Imperial College London (IUCL), UK; Universitat Pompeu Fabra (UPF), Barcelona, Spain; Inria-Asclepios project (INRIA), France. All methodologies obtained low tracking errors ($1.09mm$ to $1.52mm$ for 3DTAG, $3.51mm$ to $4.07mm$ for 3DUS, $3.09mm$ to $4.78mm$ for SSFP). Main differences lie on the amount of outliers and the smoothness of the results.

The MEVIS approach is strong due to its low computational complexity (computable in 1-2 hrs in a normal desktop). This complexity can be reduced by parallelizing the computation of each orientation. Also, since it does not require any input mask, it is applicable to any modality and target organ. This approach is weak if prior

⁹<http://www.cardiacatlas.org/>

information is desirable (i.e. tissue elasticity). Also, due to lack of temporal consistency, drift errors can be large.

The IUCL approach is strong since it complements imaging data from SSFP and 3DTAG. The required input mask is computed fully automatically. This approach is also robust to drift effects, resulting in smooth deformed meshes. However, the smoothing effect may hamper the computation of large deformations in some cases. It becomes weak in cases of spatial misalignment (i.e. different MR acquisitions), since combination relies on DICOM positioning information. The current implementation targets the integration of different MR sequences. Similar concepts could be extended to ultrasound. Radial motion is clearly assessable from parasternal US images. Combining these in the current framework or with other ultrasound acquisitions (e.g. 2D/3D apical views) could complement IUCL method or translate it to handle the US modality. In any case, the main difficulty will be the extension of the fully automated segmentation method to handle ultrasound specific challenges (reverberation artefacts, speckle, pericardium etc.).

The UPF approach is strong since it is multimodal and it guarantees temporal consistency. Similarly to IUCL, this approach is robust to drift effects, resulting in smooth deformed meshes. However, with the current implementation, the temporal smoothing is difficult to modulate. This may result in over-smoothed deformations. A main weakness of this approach is its high computational complexity. Also, it requires an input mask to define the myocardial domain.

The INRIA approach is strong since it is multimodal. It includes physiological constraints (incompressibility and elasticity) which give a clear advantage in the computation of radial deformation. A main weakness is that it creates irregularities on the deformed meshes. However, this can be easily improved by increasing the weight given to the regularization term. Similarly to UPF, this approach is computationally complex and requires an input mask to define the myocardial domain.

In future implementations, each algorithm can be fine tuned to target a specific application. For instance: to obtain smoother displacement fields (i.e. MEVIS approach); to increase/decrease the weight of SSFP and capture larger longitudinal deformations (i.e. ICUL approach); to relax the temporal smoothing term and obtain larger deformations (i.e. UPF approach); to relax incompressibility constraints in cases of reduced longitudinal deformation (i.e. INRIA approach).

A possible extension of this paper is to unify the different methodologies in a common implementation framework. The new framework should be modular in defining metrics that can take several image sequences (i.e. combine SSFP and 3DTAG). It also should be able to apply physical constraints such as incompressibility for both velocity and displacements. Provided that these features could be solved, a possible integration scheme for SSFP/3DTAG would be: 1) execute the UPF method

(as it has been reported to be robust to drift); 2) average displacement fields and execute the IUCL method; 3) from the resulting collection of displacements, execute the INRIA method to obtain an incompressible displacement field. For 3DUS, a possible integration scheme would be: 1) execute the MEVIS method algorithm to get a sparse estimate of the displacement field; 2) execute the UPF algorithm taking the MEVIS output as initialization; 3) from the resulting collection of displacements, execute the INRIA algorithm to obtain an incompressible displacement field.

Finally, this study aids the benchmarking of current cardiac motion tracking algorithms by providing a reproducible evaluation framework. As the field of cardiac motion analysis is likely to remain an open challenge for many years, both data and metrics of this study are publicly available for the continuous benchmarking of algorithms.

Acknowledgments

The work of MEVIS was partially funded by the European Regional Development Fund (ERDF). The work of IUCL was partially funded by EPSRC grant EP/H019847/1. The work of INRIA was partially funded by the Care4Me ITEA2 project. The work of UPF was partially funded by the Spanish Ministry of Innovation and Science (CEN-20091044 and TIN2009-14536-C02-01), the Catalan Institution for Research and Advanced Studies (ICREA) and the European Commission Seventh Framework Programme (FP7-ICT-2007-2-224495). The work of KCL was partially funded by FP7 euHeart project (FP7-ICT-2007-2-224495), EPSRC grant EP/H019898/1 and the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

- Amini, A., Chen, Y., Elayyadi, M., Radeva, P., 2001. Tag surface reconstruction and tracking of myocardial beads from spamm-mri with parametric b-spline surfaces. *Medical Imaging, IEEE Transactions on* 20, 94–103.
- Angelini, E., Gerard, O., 2006. Review of myocardial motion estimation methods from optical flow tracking on ultrasound data, in: *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE, IEEE*. pp. 1537–1540.
- Arsigny, V., Commowick, O., Pennec, X., Ayache, N., 2006. A log-euclidean framework for statistics on diffeomorphisms, in: *Proc. of MICCAI'06, Part I, Springer*. pp. 924–931.
- Arts, T., Prinzen, F., Delhaas, T., Milles, J., Rossi, A., Clarysse, P., 2010. Mapping displacement and deformation of the heart with local sine-wave modeling. *Medical Imaging, IEEE Transactions on* 29, 1114–1123.
- Bistoquet, A., Oshinski, J., vSkrinjar, O., 2008. Myocardial deformation recovery from cine mri using a nearly incompressible biventricular model. *Medical image analysis* 12, 69–85.
- Chandrashekhara, R., Mohiaddin, R., Rueckert, D., 2004a. Analysis of 3-d myocardial motion in tagged mr images using nonrigid image registration. *Medical Imaging, IEEE Transactions on* 23, 1245–1250.
- Chandrashekhara, R., Mohiaddin, R., Rueckert, D., 2004b. Analysis of 3-D myocardial motion in tagged MR images using nonrigid image registration. *IEEE Transactions on Medical Imaging*, 1245–1250.

- Chen, T., Wang, X., Chung, S., Metaxas, D., Axel, L., 2010. Automated 3d motion tracking using gabor filter bank, robust point matching, and deformable models. *Medical Imaging, IEEE Transactions on* 29, 1–11.
- Crosby, J., Amundsen, B., Hergum, T., Remme, E., Langeland, S., Torp, H., 2009. 3-d speckle tracking for assessment of regional left ventricular function. *Ultrasound in medicine & biology* 35, 458–471.
- De Craene, M., Piella, G., 2012. An implementation of TDFFD and LDFFD algorithms. *Insight Journal* <http://hdl.handle.net/10380/3345>.
- De Craene, M., Piella, G., Camara, O., Duchateau, N., Silva, E., Doltra, A., D’hooge, J., Brugada, J., Sitges, M., Frangi, A.F., 2012a. Temporal diffeomorphic free-form deformation: Application to motion and strain estimation from 3D echocardiography. *Med Image Anal* 16, 427–450.
- De Craene, M., Tobon-Gomez, C., Butakoff, C., Duchateau, N., Piella, G., Rhode, K.S., Frangi, A.F., 2012b. Temporal diffeomorphic free form deformation (TDFFD) applied to motion and deformation quantification of tagged mri sequences, in: Camara, O., Konukoglu, E., Pop, M., Rhode, K., Sermesant, M., Young, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*, Springer Berlin / Heidelberg. pp. 68–77.
- Deng, X., Denney, T., 2004. Three-dimensional myocardial strain reconstruction from tagged mri using a cylindrical b-spline model. *Medical Imaging, IEEE Transactions on* 23, 861–867.
- D’hooge, J., Heimdal, A., Jamal, F., Kukulski, T., Bijnens, B., Rademakers, F., Hatle, L., Suetens, P., Sutherland, G., 2000. Regional strain and strain rate measurements by cardiac ultrasound: Principles, implementation and limitations. *European Journal of Echocardiography* 1, 154–170.
- Elen, A., Choi, H., Loeckx, D., Gao, H., Claus, P., Suetens, P., Maes, F., D’hooge, J., 2008. Three-dimensional cardiac strain estimation using spatio-temporal elastic registration of ultrasound images: A feasibility study. *Medical Imaging, IEEE Transactions on* 27, 1580–1591.
- Huang, J., Abendschein, D., Davila-Roman, V., Amini, A., 1999. Spatio-temporal tracking of myocardial deformations with a 4-d b-spline model from tagged mri. *Medical Imaging, IEEE Transactions on* 18, 957–972.
- Kerwin, W., Prince, J., 1998. Cardiac material markers from tagged mr images. *Medical Image Analysis* 2, 339–353.
- Knutsson, H., Andersson, M., 2005. Morphons: Segmentation using elastic canvas and paint on priors, in: *Symposium on Image Analysis SSBA, 2005*, pp. 73–76.
- Krucker, J., LeCarpentier, G., Fowlkes, J., Carson, P., 2002. Rapid elastic image registration for 3-d ultrasound. *Medical Imaging, IEEE Transactions on* 21, 1384–1394.
- Larrabide, I., Omedas, P., Martelli, Y., Planes, X., Nieber, M., Moya, J., Butakoff, C., Sebastián, R., Camara, O., De Craene, M., Bijnens, B.H., Frangi, A.F., 2009. GIMIAS: An open source framework for efficient development of research tools and clinical prototypes, in: Ayache, N., Delingette, H., Sermesant, M. (Eds.), *Functional Imaging and Modeling of the Heart*. Springer Berlin / Heidelberg. volume 5528 of *Lecture Notes in Computer Science*, pp. 417–426.
- Ledesma-Carbayo, M., Kybic, J., Desco, M., Santos, A., Suhling, M., Hunziker, P., Unser, M., 2005. Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation. *Medical Imaging, IEEE Transactions on* 24, 1113–1126.
- Linguraru, M., Kabla, A., Vasilyev, N., del Nido, P., Howe, R., 2007. Real-time block flow tracking of atrial septal defect motion in 4d cardiac ultrasound, in: *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, IEEE. pp. 356–359.
- Mansi, T., Pennec, X., Sermesant, M., Delingette, H., Ayache, N., 2011. iLogDemos: A Demons-Based Registration Algorithm for Tracking Incompressible Elastic Biological Tissues. *Int. J. of Computer Vision*.
- Manzke, R., Lutz, A., Schenderlein, M., Bornstedt, A., Chan, R.C., Dietmeyer, K., Rasche, V., 2010. A new PVA-based dynamic cardiac phantom for evaluation of functional MR imaging methods at 3T. *ISMRM proceedings*.
- McLeod, K., Prakosa, A., Mansi, T., Sermesant, M., Pennec, X., 2012. An incompressible log-domain demons algorithm for tracking heart tissue, in: Camara, O., Konukoglu, E., Pop, M., Rhode, K., Sermesant, M., Young, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*, Springer Berlin / Heidelberg. pp. 55–67.
- Moore, C.C., Lugo-Olivieri, C.H., McVeigh, E.R., Zerhouni, E.A., 2000. Three-dimensional systolic strain patterns in the normal human left ventricle: characterization with tagged mr imaging. *Radiology* 214, 453–466.
- Papademetris, X., Sinusas, A., Dione, D., Constable, R., Duncan, J., 2002. Estimation of 3-d left ventricular deformation from medical images using biomechanical models. *Medical Imaging, IEEE Transactions on* 21, 786–800.
- Petitjean, C., Rougon, N., Cluzel, P., 2005. Assessment of myocardial function: a review of quantification methods and results using tagged mri. *Journal of Cardiovascular Magnetic Resonance* 7, 501–516.
- Petitjean, C., Rougon, N., Cluzel, P., Preteux, F., Grenier, P., 2004. Quantification of myocardial function using tagged mr and cine mr images. *The International Journal of Cardiovascular Imaging (formerly Cardiac Imaging)* 20, 497–507.
- Remme, E., Augenstein, K., Young, A., Hunter, P., 2005. Parameter distribution models for estimation of population based left ventricular deformation using sparse fiducial markers. *Medical Imaging, IEEE Transactions on* 24, 381–388.
- Rougon, N., Petitjean, C., Preteux, F., Cluzel, P., Grenier, P., 2005. A non-rigid registration approach for quantifying myocardial contraction in tagged mri using generalized information measures. *Medical Image Analysis* 9, 353–375.
- Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D., 1999. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 712–721.
- Rutz, A.K., Ryf, S., Plein, S., Boesiger, P., Kozerke, S., 2008. Accelerated whole-heart 3d cspamm for myocardial motion quantification. *Magn Reson Med* 59, 755–763.
- Shi, W., Zhuang, X., Wang, H., Duckett, S., Luong, D., Tobon-Gomez, C., Tung, K., Edwards, P., Rhode, K., Razavi, R., Ourselin, S., Rueckert, D., 2012. A comprehensive cardiac motion estimation framework using both untagged and 3-d tagged mr images based on nonrigid registration. *Medical Imaging, IEEE Transactions on* 31, 1263–1275.
- Shi, W., Zhuang, X., et al., 2011. Automatic segmentation of different pathologies from cardiac cine MRI using registration and multiple component EM estimation, in: *Functional Imaging and Modeling of the Heart*, pp. 163–170.
- Studholme, C., Hill, D., Hawkes, D., et al., 1999. An overlap invariant entropy measure of 3D medical image alignment. *Pattern recognition* 32, 71–86.
- Surry, K.J.M., Austin, H.J.B., Fenster, A., Peters, T.M., 2004. Poly(vinyl alcohol) cryogel phantoms for use in ultrasound and mr imaging. *Physics in Medicine and Biology* 49, 5529.
- Sutherland, G., Hatle, L., Claus, P., D’hooge, J., Bijnens, B., 2006. *Doppler Myocardial Imaging- A Textbook*. Hasselt, Belgium: BSWK.
- Tautz, L., Hennemuth, A., Peitgen, H.O., 2012. Motion analysis with quadrature filter based registration of tagged MRI sequences, in: Camara, O., Konukoglu, E., Pop, M., Rhode, K., Sermesant, M., Young, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*, Springer Berlin / Heidelberg. pp. 78–87.
- Tobon-Gomez, C., De Craene, M., Dahl, A., Kapetanakis, S., Carr-White, G., Lutz, A., Rasche, V., Etyngier, P., Kozerke, S., Schaeffter, T., Riccobene, C., Martelli, Y., Camara, O., Frangi, A.F., Rhode, K.S., 2012. A multimodal database for the 1st cardiac motion analysis challenge, in: Camara, O., Konukoglu, E., Pop, M., Rhode, K., Sermesant, M., Young, A. (Eds.), *Statistical Atlases*

- and Computational Models of the Heart. Imaging and Modelling Challenges, Springer Berlin / Heidelberg. pp. 33–44.
- Tustison, N., Amini, A., 2006. Biventricular myocardial strains via nonrigid registration of anfigatomical nurbs models. *Medical Imaging*, IEEE Transactions on 25, 94–112.
- Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2008. Symmetric log-domain diffeomorphic registration: A demons-based approach, in: *Proc. of MICCAI'08, Part I*, Springer, New York, USA. pp. 754–761. PMID: 18979814.
- Veress, A., Weiss, J., Klein, G., Gullberg, G., 2002. Quantification of 3d left ventricular deformation using hyperelastic warping: comparisons between mri and pet imaging, in: *Computers in Cardiology*, 2002, IEEE. pp. 709–712.
- Wandt, B., Bojo, L., Hatle, L., Wranne, B., 1998. Left ventricular contraction pattern changes with age in normal adults. *Journal of the American Society of Echocardiography* 11, 857 – 863.
- Wang, H., Amini, A., 2011. Cardiac motion and deformation recovery from mri: A review. *Medical Imaging*, IEEE Transactions on , 1–1.
- Wang, H., Shi, W., Zhuang, X., Duckett, S., Tung, K., Edwards, P., Razavi, R., Ourselin, S., Rueckert, D., 2012. Automatic cardiac motion tracking using both untagged and 3D tagged MR images, in: *Camara, O., Konukoglu, E., Pop, M., Rhode, K., Sermesant, M., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*, Springer Berlin / Heidelberg. pp. 45–54.
- Xu, C., Pilla, J., Isaac, G., Gorman, J., Blom, A., Gorman, R., Ling, Z., Dougherty, L., 2010. Deformation analysis of 3d tagged cardiac images using an optical flow method. *Journal of Cardiovascular Magnetic Resonance* 12, 1–14.
- Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D., 2008. Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. *IEEE Transactions on Medical Imaging* , 1668–1681.
- Zhuang, X., Rhode, K., Razavi, R., Hawkes, D.J., Ourselin, S., 2010. A Registration-Based Propagation Framework for Automatic Whole Heart Segmentation of Cardiac MRI. *IEEE Transactions on Medical Imaging* , 1612–1625.

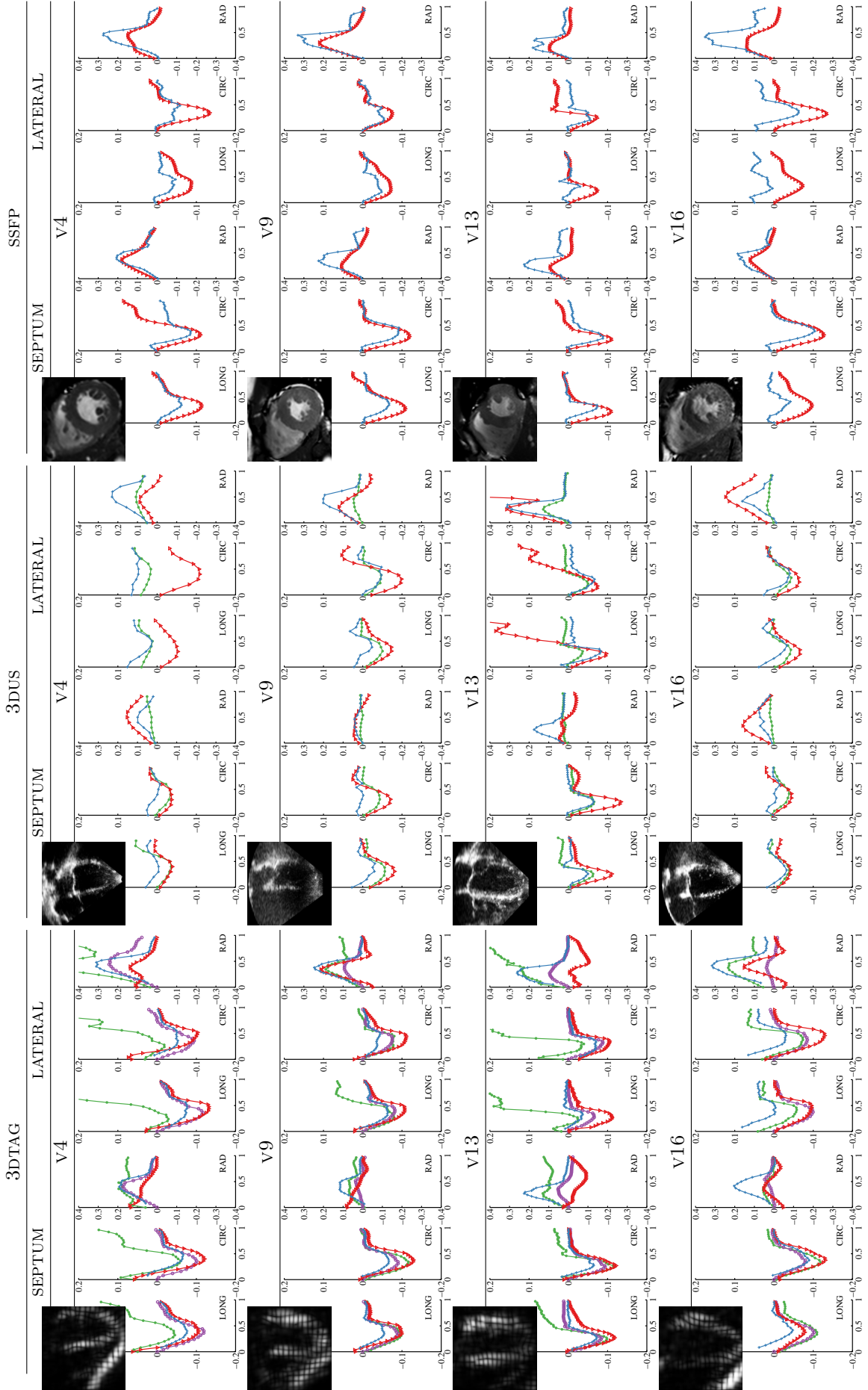


Figure 11: Longitudinal (LONG), circumferential (CIRC) and radial (RAD) strain curves obtained from 3DTAG, 3DUS and SSFP datasets. Curves were plotted as a function of time (normalized by one heart period). A vignette of each dataset at end systole is shown at the upper-left corner of the strain plots. MEVIS= green; IUCI= purple; UPF= red; INRIA= blue.